

Chapter 4

Section 4.1

Check Your Understanding, page 213:

1. The company inspector is using a convenience sample. This could lead him to overestimate the quality of the oranges if the farmer puts the best oranges on top or if the oranges at the bottom of the crate are damaged from the weight of oranges on top of them.
2. *Nightline* was using a voluntary response sample. Only those who feel particularly strongly about the issue are likely to respond. In this case, those who are happy that the United Nations has its headquarters in the U.S. already have what they want and so are less likely to worry about responding to the question. This means that the proportion who answered “No” in the sample is likely to be higher than the true proportion in the U.S. who would answer “No.”

Check Your Understanding, page 223:

1. It might be difficult to give a survey to an SRS of 200 fans because you would have to identify 200 different seats, go to those seats in the arena and find the people who are sitting there. This means going to 200 different locations throughout the arena, which would take time. There is also the problem that people are not always in their seats throughout the game and not all seats will necessarily be occupied in any given game.
2. For a stratified sample, it is best to create strata where the people within a stratum are very similar to each other but different than the people in other strata. In this case, it would be better to take the lettered rows as the strata because each lettered row is the same distance from the court and so would contain only seats with the same (or nearly the same) ticket price. This means that all people in any given stratum would have paid roughly the same amount for their tickets.
3. For a cluster sample, it is best if the people in each cluster reflect the variability found in the population. In this case, it would be better to take the numbered sections as the clusters because they include all different seat prices. Each section contains seats with many different ticket prices so the people in a section would mirror the characteristics of the population as a whole.

Check Your Understanding, page 228:

1. (a) Using the telephone directory will result in undercoverage because those who are not listed in the phone book (those who do not have a phone or have only a cell phone) do not have the opportunity to be chosen.
(b) If a person that is selected for the sample cannot be contacted, it is nonresponse.
(c) Using a convenience sample leads to undercoverage, because people who are not walking on that street do not have the opportunity to be chosen.
2. By making it sound like they are not a problem in the landfill, this question will result in fewer people suggesting that we should ban disposable diapers. The proportion who would say “Yes” to this survey question is likely to be smaller than the proportion who would say “Yes” to a more fairly-worded question.

Exercises, page 229:

- 4.1 The population is all local businesses. The sample is the 73 businesses that return the questionnaire
- 4.2 The population is all the artifacts discovered at the dig. The sample is those artifacts (2% of the population) that are chosen for inspection.

4.3 The population is the 1000 envelopes stuffed during a given hour. The sample is the 40 randomly selected envelopes.

4.4 The population is all 45,000 people who made credit card purchases. The sample is the 137 people who returned the survey form.

4.5 This is a voluntary response sample and only represents the opinions of those who feel very strongly about this issue. In this case, it appears that people who strongly support gun control volunteered more often, causing the proportion in the sample to be greater than the proportion in the population.

4.6 Letters to legislators are an example of a voluntary response sample—the proportion of letters opposed to the insurance should not be assumed to be a fair representation of the attitudes of the congresswoman’s constituents. Only those who have very strong opinions will write in. In this case, it is likely that the true proportion of constituents who oppose the bill is less than 871/1128.

4.7 This is a voluntary response sample and overrepresents the opinions of those who feel most strongly about the issue being surveyed.

4.8 This is a convenience sample. The sample is likely to overestimate the unemployment rate because people without jobs have more time to be at the mall than those who are employed.

4.9 (a) A convenience sample.

(b) It is unlikely that the first 100 students to arrive at school are representative of the student population in general. 7.2 hours is probably less than the true average because students who arrive first to school had to wake up earlier and may have gotten less sleep than those students who are able to sleep in.

4.10 (a) A voluntary response sample.

(b) It is biased toward readers who feel most strongly about the issue. The reported value of 85% is probably higher than the true percent. Readers who have been involved in an accident caused by cell phone use are more likely to respond to the poll and say “Yes.”

4.11 (a) Number the 40 students from 01 to 40 alphabetically. Go to the random number table and pick a starting point. Record two-digit numbers, skipping any numbers that aren’t between 01 and 40 and any repeated numbers, until you have 5 unique numbers between 01 and 40. Use the 5 students corresponding to these numbers.

(b) Starting at line 107 we read off the following numbers: 82 (ignore) 73 (ignore) 95 (ignore) 78 (ignore) 90 (ignore) **20** (select) 80(ignore) 74 (ignore) 75 (ignore) **11**(select) 81 (ignore) 67 (ignore) 65 (ignore) 53 (ignore) 00 (ignore) 94 (ignore) **38** (select) **31** (select) 48 (ignore) 93 (ignore) 60(ignore) 94 (ignore) **07** (select). We have selected: Johnson (20), Drasin (11), Washburn (38), Rider (31), and Calloway (07).

4.12 (a) Number the 33 complexes from 01 to 33 alphabetically. Go to the random number table and pick a starting point. Record two-digit numbers, skipping any that aren’t between 01 and 33 and any repeated numbers, until you have 3 unique numbers between 01 and 33. Use the 3 complexes corresponding to these numbers.

(b) Starting at line 117 we read off the following numbers: 38 (ignore) **16** (select) 79 (ignore) 85 (ignore) **32** (select) 62 (ignore) **18** (select). We have selected: Fairington (16), Waterford Court (32) and Fowler (18).

4.13 (a) *Using calculator*: Number the plots from 1 to 1410. Use the command `randInt(1,1410)` to select 141 different integers from 1 to 1410 and use the corresponding 141 plots. *Using Table D*: Number the plots from 0001 to 1410. Go to the random number table and pick a starting point. Record four-digit numbers, skipping any that aren't between 0001 and 1410 and any repeated numbers, until you have 141 unique numbers between 0001 and 1410. Use the corresponding 141 plots.

(b) *Using calculator*: Answers will vary. *Using Table D*: Starting at line 131 we read off the following numbers: **0500** 7166 3281 1941 4873 **0419** 7855 7645 1959 6565 6873 2552 5984 2920 8796 4316 5937 3931 6859 7150 4574 **0418**. The first three plots in our sample are plots 0500, 0419 and 0418.

4.14 (a) *Using calculator*: Number the gravestones from 1 to 55,914. Use the command `randInt(1,55914)` to select 395 different integers from 1 to 55,914 and use the corresponding 395 plots. *Using Table D*: Number the gravestones from 00001 to 55914. Go to the random number table and pick a starting point. Record 5-digit numbers, skipping any that aren't between 00001 and 55914 and any repeated numbers, until you have 395 unique numbers between 00001 and 55914. Use the corresponding 395 plots.

(b) *Using calculator*: Answers will vary. *Using Table D*: Starting at line 127 we read off the following numbers: **43909** 99477 **25330** 64359 **40085**. The first three gravestones selected are 43909, 25330, and 40085.

4.15 (a) False—although on average there will be four 0s in every set of 40 digits, the number of 0s can be less than 4 or greater than 4 by chance.

(b) True—there are 100 pairs of digits 00 through 99, and all are equally likely.

(c) False—0000 is just as likely as any other string of four digits.

4.16 If you always begin at the same place, then the results would not be random. You would end up using the same sample in every case.

4.17 (a) It might be difficult to locate the 20 phones from among the 1000 produced that day. Also, assuming none of the phones can be shipped until after the inspection, inspecting a random sample of 20 phones could hold up the shipping process.

(b) It is possible that the quality of the phones produced changes over the course of the day so that the last phones manufactured are not representative of the day's production.

(c) This is not an SRS because each sample of 20 phones does not have the same probability of being selected. In an SRS, it is possible for 2 consecutive phones to be selected in a sample, but this is not possible with a systematic random sample.

4.18 (a) To obtain an SRS, every tree would need to be identified and numbered. It is not practical to even identify every tree in the park.

(b) This convenience sampling method is biased because these trees are unlikely to be representative of the population. Trees along the main road are more likely to be damaged by cars and people, and may be more susceptible to infestation.

(c) The scientists can be confident that the percentage of all pine trees in the area that are infected by the pine beetle is near 35%, however the percentage is unlikely to be exactly 35% because of sampling variability.

4.19 Assign numbers 01 to 30 to the students. Pick a starting point on the random digit table. Record two-digit numbers, skipping any that aren't between 01 and 30 and any repeated numbers, until you have 4 unique numbers between 01 and 30. Use the corresponding four students. Then, assign numbers 0 to 9 to the faculty members. Continuing on the table, record one-digit numbers, skipping any repeated numbers, until you have 2 unique numbers between 0 and 9. Use the corresponding faculty members. Starting on line 123 gives 08-Ghosh, 15-Jones, 07-Fisher, and 27-Shaw for the students and 1-Besicovitch and 0-Andrews for the faculty.

4.20 Label the 500 midsize accounts with numbers from 001 to 500. Pick a starting point on the random digit table. Record three-digit numbers, skipping any that aren't between 001 and 500 and any repeated numbers, until you have 25 unique numbers between 001 and 500. Verify the corresponding 25 accounts. Then, label the 4400 small accounts with numbers from 0001 to 4400. Continuing on the random digit table, record four-digit numbers, skipping any that aren't between 0001 and 4400 and any repeated numbers, until you have 44 unique numbers between 0001 and 4400. Verify the corresponding 44 accounts. Starting at line 115, the first three midsize accounts are 417, 494, and 322. The first three small accounts are 2470, 1893, and 3259.

4.21 (a) Use the three types of seats (sideline, corner, and end zone) as the three strata because ticket prices will be similar within each stratum but different between the three strata. People who can afford more expensive tickets probably have different opinions about the concessions than people who can only afford the cheaper tickets.

(b) A stratified random sample will include seats from all over the stadium, which would make it very time consuming to get to everyone. A cluster sample would be easier to obtain, because there would be many people sitting all together who would be part of the sample. In this case, it would be easiest to use each numbered section as a cluster.

4.22 (a) Because satisfaction with the property is likely to vary depending on the location of the room, we should stratify by floor and view. For example, one stratum would be rooms on the first floor with a water view and another stratum would be rooms on the 30th floor with a golf course view. Using a stratified random sample would assure the manager that he got opinions from each type of room and provide a more precise estimate of customer satisfaction.

(b) We could use floors as clusters. This would be a simpler option because the manager would only need to survey guests on three floors instead of having to survey guests all over the hotel.

4.23 No. In an SRS, each possible sample of 250 engineers is equally likely to be selected. However, the method described restricts the sample to having exactly 200 males and 50 females. For example, a sample of 199 males and 49 females cannot be selected with this method but could be selected in an SRS.

4.24 In an SRS, each possible sample of 5 students is equally likely to be selected. However, the method described restricts the sample to having exactly 3 students over 21 and 2 students under 21. For example, a sample of 4 students over 21 and 1 student under 21 cannot be selected with this method but could be selected in an SRS.

4.25 (a) This is cluster sampling.

(b) The cable company could have chosen this method to save time and money. In an SRS, the company would have to visit individual homes all over the rural subdivision. With the cluster sampling method, the company only has to visit 5 locations.

4.26 (a) This is cluster sampling.

(b) The lumber company could have chosen this method to save time and money. In an SRS, the company would have to inspect trees all over the entire forest. With the cluster sampling method, the company only has to visit 20 locations.

4.27 (a) Because different random samples will include different students and produce different estimates, it is unlikely that the sample result will be the same as the proportion of all students at the school who use Twitter.

(b) An SRS of 100 students is more likely to get a sample result close to the true population value. Larger random samples give us better information about the population than smaller random samples.

4.28 (a) Because different random samples will include different students and produce different estimates, it is unlikely that the sample result will be the same as the mean distance that all students live from campus.

(b) An SRS of 100 students is more likely to get a sample result close to the true population value. Larger random samples give us better information about the population than smaller random samples.

4.29 This survey will yield a biased result because you are sampling only from the lower priced ticket holders. This will likely produce an estimate that is too small, as fans in the club seats and box seats probably spend more money at the game than fans in cheaper seats.

4.30 This survey will yield a biased result because administrators are only sampling from students who take the bus and not including students who drive themselves to school or who are dropped off. This will likely produce an estimate that is too large, as students who take the bus probably need to wake up earlier so they don't miss the bus.

4.31 (a) The response rate was $\frac{5,029}{45,956} = 0.1094$ so the nonresponse rate is $1 - 0.1094 = 0.8906$ or 89.1%.

(b) This survey will yield a biased result because the people who have long commutes are less likely to be at home and be included in the sample. This will likely produce an estimate that is too small, as the people who are at home to answer the survey probably have shorter commutes (or none at all).

4.32 The higher no-answer rate was probably the second period—when families are likely to be vacationing or spending time outdoors. A high rate of nonresponse makes sample results less reliable because you don't know how these individuals would have responded. It is very risky to assume that they would have responded exactly the same way as those individuals who did respond.

4.33 We would not expect very many people to claim they have run red lights when they haven't, but some people will deny running red lights when they have. Thus, we expect that the sample proportion underestimates the true proportion of drivers who have run a red light.

4.34 People likely claim to wear their seat belts because they know they should; they are embarrassed or ashamed to say that they do not always wear seat belts. Such bias is likely in most surveys about seat belt use (and similar topics).

4.35 (a) The wording is clear, but the question is slanted in favor of warning labels because of the first sentence stating that some cell phone users have developed brain cancer.

(b) The question is clear, but it is slanted in favor of national health insurance by asserting it would reduce administrative costs and not providing any counter-arguments. The phrase "do you agree" also pushes respondents toward the desired response.

(c) The wording is too technical for many people to understand. For those who do understand the question, it is slanted because it suggests reasons why one should support recycling. It could be rewritten to something like: "Do you support economic incentives to promote recycling?"

4.36 (a) The question is clear, but the two options presented are too extreme; no middle position on gun control is allowed. Also, the wording pushes respondents to choose option 2. The language used in option 2 is from the Constitution and people might avoid option 1 because they don't like the idea of government confiscating personal property.

(b) The question is so complicated that it isn't clear. The phrasing of this question will tend to make people respond in favor of a nuclear freeze because only one side of the issue is presented.

4.37 c

4.38 c

4.39 d

4.40 e

4.41 d

4.42 c

4.43 (a) For each additional day, the predicted sleep debt increases by about 3.17 hours.

(b) Although the slope of 3.17 is pretty close to the 3 hours/day of sleep debt claimed by the researcher, the predicted sleep debt for a 5-day school week, based on the least-squares regression equation, is $2.23 + 3.17(5) = 18.08$ hours. This is about 3 hours more than the researcher claimed for a 5-day week, so the students have reason to be skeptical of the research study's reported results.

4.44 (a) If a bandwidth measurement is at the 95th percentile, then 95% of the other bandwidth measurements will be less than this amount.

(b) The method using the 98th percentile would cost the company more because it would suggest a higher usage of bandwidth by the company.